Published online 30 October 2012 *Nucleic Acids Research, 2013, Vol. 41, Database issue* ***D125–D131***
doi:10.1093/nar/gks997

SpliceAid-F: a database of human splicing factors and their RNA-binding sites

Matteo Giulietti[1,2], Francesco Piva[2], Mattia D'Antonio[1,3], Paolo D'Onorio De Meo[3], Daniele Paoletti[3], Tiziana Castrignanò[3], Anna Maria D'Erchia[1,4], Ernesto Picardi[1,4], Federico Zambelli[5], Giovanni Principato[2], Giulio Pavesi[5] and Graziano Pesole[1,4,6,*]

[1]Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Bari, [2]Department of Specialized Clinical Sciences and Odontostomatology, Polytechnic University of Marche, Ancona, [3]Consorzio per le Applicazioni di Supercalcolo per Università e Ricerca, Rome, [4]Institute of Biomembranes and Bioenergetics, Consiglio Nazionale delle Ricerche Bari, [5]Department of Biosciences, University of Milan, Milan and [6]Center of Excellence in Genomics (CEGBA), Bari, Italy

Received August 8, 2012; Revised September 20, 2012; Accepted September 30, 2012

ABSTRACT

A comprehensive knowledge of all the factors involved in splicing, both proteins and RNAs, and of their interaction network is crucial for reaching a better understanding of this process and its functions. A large part of relevant information is buried in the literature or collected in various different databases. By hand-curated screenings of literature and databases, we retrieved experimentally validated data on 71 human RNA-binding splicing regulatory proteins and organized them into a database called 'SpliceAid-F' (http://www.caspur.it/SpliceAidF/). For each splicing factor (SF), the database reports its functional domains, its protein and chemical interactors and its expression data. Furthermore, we collected experimentally validated RNA–SF interactions, including relevant information on the RNA-binding sites, such as the genes where these sites lie, their genomic coordinates, the splicing effects, the experimental procedures used, as well as the corresponding bibliographic references. We also collected information from experiments showing no RNA–SF binding, at least in the assayed conditions.

In total, SpliceAid-F contains 4227 interactions, 2590 RNA-binding sites and 1141 'no-binding' sites, including information on cellular contexts and conditions where binding was tested.

The data collected in SpliceAid-F can provide significant information to explain an observed splicing pattern as well as the effect of mutations in functional regulatory elements.

INTRODUCTION

Pre-mRNA splicing is an essential nuclear process where intronic sequences are identified and removed, joining the flanking exons to generate mature mRNAs, which can be then translocated and translated in the cytoplasm. The splicing reaction, which has been recently shown to be also affected by the chromatin status (1), is catalyzed by the spliceosome, a large (~60S) RNA–protein molecule (2) that interacts with RNA elements like the 5′ and 3′ splice sites (ss) defining the exon/intron boundaries, the polypyrimidine tract that aids in 3′ss recognition and the branch-point sequence.

Indeed, in pre-mRNAs, there might be many other elements helping the definition of exon/intron boundaries, but it is not yet fully known how the splicing machinery selects the 5′ and 3′ss to be used. It has been shown how the choice seems to be driven by other *cis*-acting elements (3), which can be located within both exons and introns. These regulatory sequences can be bound by many different splicing regulatory proteins or "splicing factors" (SFs), interacting with each other and leading to the precise definition of a sequence as exon or intron (4). In other words, SFs decode the information in the pre-mRNA sequence by recognizing RNA-binding motifs. Mutagenesis experiments have highlighted that splicing regulatory elements are scattered within the entire pre-mRNA sequence, and any transcribed nucleotide could be potentially involved in the generation of the splicing patterns.

Nearly all human transcripts can yield several different splicing isoforms owing to the competition among different factors for the same binding sites. Moreover, some alternative splicing forms are cell-specific, yielding different protein isoforms in different cells (5–8). Therefore, it is evident that knowledge of the RNA-binding specificity of

*To whom correspondence should be addressed. Tel: +39 080 5443588; Fax: +39 080 5443317. Email: graziano.pesole@biologia.uniba.it

the *trans*-acting SFs is essential for predicting their binding activity and to fully understand splicing regulatory mechanisms.

The known binding sites of SFs are usually summarized as consensus target sites, defined by the most frequent nucleotide found at each position of the sites. This description has clear limitations, and tools using this model (9–12) yield poor binding position predictions (13). To overcome this problem and reduce the false-positive results, descriptors should be built taking advantage of information from the full set of original binding sites, and if possible also from sites experimentally shown not to be bound by the SF.

An SF 'no-binding' site is defined as a sequence element for which the SF has been experimentally shown to have no binding affinity in one or more cellular contexts. In contrast, an SF-binding site is a sequence element that specifically binds the SF in at least one condition (e.g. cell type, physiological status, etc.). Another useful piece of information is then the binding specificity in different cellular contexts (conditional-binding sites, Table 3), as it is possible that the same motif may be bound or not by the same SF in different conditions. A curated data collection of validated binding and no-binding sites for a given factor, complete with their cellular specificity, can be thus of relevant interest, for example, to train a machine learning algorithm to accurately discriminate and predict binding sites for a given SF, or to assess which sequence elements seem to be more important for SF binding. Unfortunately, few data on binding sites and their specificity are as yet available, whereas data on no-binding sites have never been collected.

To better understand the splicing process and function, it is important to shed light on the complex network of interactions involving SFs and their RNA target sites. This goal is made difficult by the high number of RNA-binding proteins, some of them still unknown, the limited information about their functional features and the fact that available information is fragmented in a large number of papers. Therefore, it would be useful to collect in a unique resource all available information about SFs.

We previously introduced two databases, SpliceAid (14) and SpliceAid 2 (15), consisting of RNA motifs bound by SFs, together with an interface designed to detect the presence of putative motifs in sequences submitted by users.

Here we present SpliceAid-F, an exhaustive repository of SFs including information on their expression, functional domains, protein and chemical interactors and RNA-binding and no-binding sites. Information of binding sites includes the corresponding genes, their genomic coordinates, the splicing effect, the experimental procedures used to assess binding and the related references. All these data have been manually extracted and collected by extensive database and literature screenings, yielding information on 71 splicing proteins, all those investigated so far to our knowledge for which some information is available on their RNA targets.

DATABASE DESIGN AND IMPLEMENTATION

Database structure description

SpliceAid-F contains information on human SFs, organized in a relational database structured in four tables. In the main table, SFs are listed together with data concerning both the gene encoding the SF protein and the protein itself. In particular, we collected official gene symbols, their synonyms, NCBI Gene IDs and their links to NCBI HomoloGene. Regarding proteins, the databases lists the UniProt IDs, the type of RNA-binding domains and the known interacting proteins according to IntAct (16), STRING (17), DIP (18), MINT (19), HPRD (7) and BioGRID (20) databases and the interacting chemicals according to IntAct (16), STITCH (17) and ProtChemSI (21). Also, the interface reports the diseases associated with the alterations of SF expression or mutations in the encoding gene (provided by OMIM database), links to expression databases (Cancer Genome Anatomy Project, Human Protein Atlas, Human Protein Reference Database, Human Proteinpedia) and CLIP (Crosslinking and ImmunoPrecipitation) data.

Each record in the main table corresponds to an SF, and it is linked to records of the second and third tables, collecting more data. In particular, the second table contains the known RNA-binding sites for the SF, while the third lists the SF no-binding sites. We also collected experimental information on human SF binding to RNAs from other species (listed in Supplementary Table S1), including viruses using splicing machinery of the host cell.

For each binding site, we collected the species it belongs to, its genomic coordinates and the corresponding gene. The database also collects the experimental procedures used to assess the binding and the cell or tissue investigated, the splicing activity (in terms of exon definition or intron definition) of the SF and the diseases possibly associated with a binding site alteration (provided by OMIM). Finally, the fourth table of database stores, for each SF-binding or no-binding site, the related literature references.

Protein annotation

Overall, the database contains data on 71 human SFs, including 13 SR proteins, 27 heterogeneous nuclear ribonucleoprotein (hnRNP) proteins and other 31 proteins belonging to the ELAV, KHDRBS, CELF, Nova and Fox families. SR and hnRNP proteins mainly host one or more RNA recognition motif (RRM) domains, whereas other splicing proteins contain different RNA-binding domains other than RRM, such as K-homology domains, zinc fingers and others (Table 1).

Concerning chemical interactions, we performed a hand-curated filtering of the data to include in the database only the true interactors for each of the 71 splicing proteins. In fact, many interactions reported in STITCH and IntAct seemed to be inaccurate. For example, CUGBP1 (22), FMRP (23), hnRNP E2 (24–26) and SAP155 (27) do not interact with selenomethionine, but this modified amino acid is incorporated by researchers instead of methionine to aid the structure

Table 1. RNA-binding proteins collected in SpliceAid-F, including information on their functional domains

Proteins		RNA-binding domain			
Class	Name	RRM	KH	Zinc finger	Other
SRs	SC35, SRp20, SRp38, SRp54, HTra2α, HTra2β1	1			
	9G8	1		1 CCHC	
	SF2/ASF, SRp30c, SRp40, SRp55, SRp75	2			
	SRm160				1 PWI
hnRNPs	hnRNP C, C1, C2, G	1			
	hnRNP P	1		1 RanBP2	
	hnRNP A0, A1, A2/B1, A3, D, D0, DL, H3	2			
	hnRNP F, H1, H2, L, LL, M, Q	3			
	hnRNP I (PTB), nPTB	4			
	hnRNP E1, E2, J, K		3		
	hnRNP U				1 SAP
Others	Fox-1, Fox-2	1			
	RBM25	1		1 PWI	
	DAZAP1, PSF, TDP43	2			
	RBM4	2		1 CCHC	
	RBM5	2		1 C2H2, 1 RanBP2	1 G-patch
	CUG-BP1, ESRP1, ESRP2, ETR-3, HuB, HuC, HuD, HuR, TIA-1, TIAL1	3			
	QKI, Sam68, SLM-1, SLM-2		1		
	SF1		1	1 CCHC	
	FMRP		2		
	Nova-1, Nova-2		3		
	KSRP		4		
	ZRANB2			2 RanBP2	
	MBNL1			4 C3H1	
	YB-1				1 CSD
	SAP155				11 HEAT repeats

RRM, RNA Recognition Motif; KH, K homology; CCHC, Cys-Cys-His-Cys; RanBP2, Ran-binding protein 2; PWI, Pro-Trp-Ile; C2H2, Cys-Cys-His-His; G-patch, Glycine rich; SAP, SAF-A/B, Acinus and PIAS; C3H1, Cys-Cys-Cys-His; HEAT, Huntingtin, EF3, PP2A, TOR1; CSD, cold-shock domain.

elucidation of proteins by X-ray crystallography. FMRP does not interact with magnesium, which is instead used in the crystallization buffer (23). Analogously, we have performed a literature screening to annotate only direct protein–protein interactions. Also in this case, we found several inaccuracies. For example, the interaction between ZRANB2 and USP39, annotated by IntAct and BioGRID, is not correct because the cited article (28) shows instead the interaction between ZRANB1 and USP39. A cautionary note is necessary for RNA binding sited detected by Cross Linking and ImmunoPrecipitation followed by high-throughput sequencing (CLIP-Seq) as this technique may also support several false positive binding sites (29).

RNA interaction annotations

RNA-binding data in SpliceAid-F is partly derived from our previous SpliceAid database (14) and partly retrieved from literature. For all binding and no-binding evidence found in literature, we report the assays and the conditions used by the authors. It should be taken into account that a no-binding observation for a particular SF refers to a specific context; therefore, the SF may bind the same element in a different condition.

Different experimental techniques have been used for the binding assessment, like EMSA, UV cross-linking and immunoprecipitation, protein affinity purification, pull-down assay, complementation assay, SELEX, filter-binding assay, homopolymer-binding assay, nuclear magnetic resonance methods, fluorescence methods, yeast three-hybrid assay and biosensor analysis. These pieces of information allow the user to evaluate the reliability of each RNA–protein interaction collected in SpliceAid-F. For example, the user may consider not-reliable-enough binding observations assessed only by EMSA or by filter-binding assays with purified proteins, as these are *in vitro* assays. In contrast, the binding might be considered to be detected in a more reliable way when UV cross-linking followed by immunoprecipitation or additional assays, as small interfering RNA (siRNA) silencing or complementation assay, is performed.

For multiple-sequence experiments, as SELEX, we considered each identified binding sequence and not the consensus motif deduced by the authors. However, because these sequences are usually much longer than the real binding site, for each of these, we report only the motif that the authors showed to be functional. This is usually performed by identifying the functional motifs through alignment of the binding sequences, followed by further assays such as, for example, mutagenesis and immunoblotting. Similarly, in case of single-sequence experiments, we followed the author's assessments about the position and the length of the functional motif to be annotated.

Moreover, we associated with each binding site a normalized score, derived by the binding affinity deduced from the experiments. The score is computed according to the criteria we previously introduced (14). Briefly, a positive score (+1 . . . +10) is assigned to target

sequences that facilitate exon definition, like exonic splicing enhancer and intronic splicing silencer motifs. Following the same criteria, we similarly assigned a negative score ($-1 \ldots -10$) to the target sequences that facilitate intron definition, that is exonic splicing silencer and intronic splicing enhancer motifs. High absolute scores correspond to high RNA-binding affinity, and hence sequence elements with a score of +10 and −10 show the greatest affinity for the SF facilitating exon and intron definition, respectively. Obviously, our scoring system does not represent accurately the binding site affinity because there are several factors, other than sequence, influencing the RNA–protein binding, like RNA secondary structure, competition or cooperation of other proteins and other cellular physical–chemical conditions. However, the score could provide an indication on how the context affects the binding when a motif binds the same factor with different scores in different contexts. The score might be also useful to train prediction algorithms or to explain the ratio among splicing isoforms as outcome of a binding competition between SFs.

We included in the database information on the splicing activity (exon definition or intron definition) of the bound SF for each RNA-binding site, as some SFs promote exon inclusion, some promote exon skipping and others promote both the events according to the genetic or cellular context. Table 2 shows our data about SF activities. In particular, SR proteins are known to possess exon definition activity, but our data show that in rare cases, they can promote exon skipping. Even SRp54 seems to have only intron definition activity, although this is claimed on the basis of only one piece of evidence. Analogously, hnRNPs have intron definition activity, but in rare cases, they can also exhibit exon definition activity.

The distribution of the collected binding and no-binding sites (Table 3) highlights the fact that for some factors, only few binding sites are available. For example, for ESRP1, ESRP2, hnRNP A0, hnRNP D0, hnRNP G, hnRNP J, hnRNP M, QKI, RBM25, SAP155, SLM-1, SRm160 and SRp54, only one binding site is known. In this case, it is unfeasible to build a model of binding preference to be used in binding site predictions. Similarly, the lack of conditional- and no-binding motifs makes hard the definition of models with adequate specificity.

Furthermore, starting from their genomic coordinates, we connected SpliceAid-F binding sites to the 'Transcript View' section of the ASPicDB database (30,31) to show where the motifs lie and to correlate them with the presence of other regulatory elements and exon/intron boundaries. For example, the presence of a binding site for a splicing silencer in an exon can explain exon skipping in some transcripts of ASPicDB.

Similar databases

We previously introduced SpliceAid (14) and SpliceAid 2 (15), database-driven tools for the prediction of splicing binding sites in user-submitted sequences by exact matching search. These tools are being used by researchers studying functional effects of gene polymorphisms (32) or mutations (33), as they predict whether these variations can affect the splicing process by altering the binding between SFs and pre-mRNA. SpliceAid-F is instead a different resource, consisting of a database collecting information about SFs, such as their expression, their orthologous genes and proteins, the associated diseases, interacting proteins and chemicals, RNA-binding domains and interacting and not-interacting RNAs. This information repository has been devised for researchers who study the splicing machinery and its functions, and bioinformaticians who want to develop new and more accurate binding site prediction tools.

Another repository of information on the activity of RNA-binding proteins is the RNA-Binding Protein DataBase (RBPDB) (34) (http://rbpdb.ccbr.utoronto.ca/index.php). Like RBPDB, SpliceAid-F collects data about RNA-binding proteins and their protein domains, but the latter is specifically designed for human SFs. Moreover, SpliceAid-F lists many other types of information about splicing proteins. We have collected only experimental sites, whereas RBPDB does not distinguish single motifs from consensus sequences. SpliceAid-F collects all known binding sites for each SF, their genomic coordinates, the corresponding gene region and its splicing activity. SpliceAid-F contains about four times more splicing RNA target sites than RBPDB (i.e. 2590 in SpliceAid-F v1.0 versus 606 in RBPDB v2.1). Finally, experimentally validated no-binding sites are only collected in SpliceAid-F.

Table 2. The activity of some splicing proteins is context dependent

Activity	SR proteins	hnRNP proteins	Other proteins
Exon definition	9G8, HTra2α, HTra2β1, SRm160, SRp55, SRp75		ESRP1, ESRP2, Nova-1, Nova-2, SLM-1, SLM-2
Mainly exon definition	SC35, SF2/ASF, SRp20, SRp30c, SRp38, SRp40		ETR-3, FMRP, RBM4, Sam68, TIA-1, TIAL1, YB-1
Mainly intron definition		hnRNP F, H1, H2, H3, I (PTB), L, Q	Fox-1, Fox-2, MBNL1
Intron definition	SRp54	hnRNP A0, A1, A2/B1, A3, C, C1, C2, D, D0, DL, E1, E2, G, J, K, LL, M, P, U, nPTB	CUG-BP1, DAZAP1, HuB, HuC, HuD, HuR, KSRP, PSF, QKI, RBM25, RBM5, SAP155, SF1, TDP43, ZRANB2

Exon definition: the proteins of this class have been shown so far to contribute exclusively to exon retention in mature transcripts. Mainly exon definition: usually these proteins promote exon retention, but in rare contexts might promote exon skipping. Mainly intron definition: proteins mostly causing exon skipping. Intron definition: proteins causing exon skipping in all contexts found.

Table 3. Distribution of the sites collected in SpliceAid-F, reporting the number of binding sites, including conditional- and no-binding sites, which have been shown unable to bind a splicing factor in some or any conditions, respectively

Splicing factor	Binding sites	Conditional-binding sites	No-binding sites
9G8	70	1	29
CUG-BP1	42	3	32
DAZAP1	12	0	3
ESRP1	1	0	1
ESRP2	1	0	1
ETR-3	31	4	36
FMRP	43	4	4
Fox-1	12	0	2
Fox-2	13	0	3
hnRNP A0	1	0	0
hnRNP A1	143	17	39
hnRNP A2/B1	42	1	8
hnRNP A3	2	0	1
hnRNP C	21	7	13
hnRNP C1	11	2	19
hnRNP C2	10	0	16
hnRNP D	29	2	14
hnRNP D0	1	0	0
hnRNP DL	34	0	0
hnRNP E1	43	17	14
hnRNP E2	39	13	23
hnRNP F	67	8	26
hnRNP G	1	0	0
hnRNP H1	85	8	45
hnRNP H2	101	8	42
hnRNP H3	60	8	44
hnRNP I (PTB)	129	13	53
hnRNP J	1	0	12
hnRNP K	58	15	13
hnRNP L	172	4	9
hnRNP LL	13	0	2
hnRNP M	1	0	3
hnRNP P (TLS)	16	4	8
hnRNP Q	10	0	7
hnRNP U	19	3	0
HTra2α	7	0	3
HTra2β1	20	1	14
HuB	44	0	1
HuC	2	0	2
HuD	51	6	5
HuR	72	25	26
KSRP	22	2	7
MBNL1	92	11	34
Nova-1	25	4	18
Nova-2	12	4	9
nPTB	3	0	1
PSF	32	0	7
QKI	1	0	0
RBM25	1	0	1
RBM4	8	0	2
RBM5	7	0	2
Sam68	16	0	5
SAP155	1	0	0
SC35	172	5	47
SF1	24	1	5
SF2/ASF	248	15	52
SLM-1	1	0	0
SLM-2	6	0	0
SRm160	1	0	0
SRp20	74	0	23
SRp30c	25	9	6
SRp38	10	0	0
SRp40	68	7	27
SRp54	1	0	0
SRp55	64	7	24
SRp75	8	0	18
TDP43	22	1	8
TIA-1	39	2	7
TIAL1	37	2	2
YB-1	21	1	14
ZRANB2	19	0	4
Total	2590	245	896

USING SpliceAid-F

The user can download the entire database from the Home page or query the database by the Search interface.

Search interface

The user can search information about a specific SF by selecting its protein name in the field 'Splicing Factor', its encoding-gene name or symbol by the field 'Gene', its target genes, its splicing activity, the chromosome or the specific gene region where the binding site lies and the site length. Moreover, users can search heterologous target sequences bound from human SFs by selecting the organism name from the 'SF binding site in' list. Also, the search criteria can be combined.

SF results

After search, the result page lists the SFs that comply with the request. For each SF, we give back the official name and synonyms of its encoding gene, the orthologous genes and proteins, the UniProt ID, the type of RNA-binding domain(s), a list of interacting proteins and chemicals, links to expression data, the disease associated with the SF alteration (if available) and links to CLIP-Seq data.

Binding site and no-binding site results

Binding data are as follows: experimentally assessed binding and no-binding sites and other experimental information such as name, ID and region of the gene where the site lies, coordinates of the target site, the type of the performed binding assay and splicing assay, the cell or tissue context, the related reference (PubMed ID), the publication year and the link to the ASPicDB. Because the splicing activity of a factor depends also on the gene context to which the target site belongs, for each binding site, we report also the specific splicing activity in terms of exon definition or intron definition. Other columns show the known diseases associated with a binding site alteration, a score meaning the binding strength.

The user can export the results in a tab-delimited plain text file or further filter the results.

CONCLUSIONS

SpliceAid-F collects into a unique resource several heterogeneous information about splicing regulatory proteins, their binding sites and context-specific activity. To our knowledge, it is the only database collecting also no-binding sites. Because SpliceAid-F is the most specific and comprehensive database on human RNA splicing, it may be a reference tool for researchers studying splicing or mutations affecting splicing. In particular, our data may help to explain the observed splicing patterns as well as the effect of mutations, as mutations located in functional regulatory elements may generate aberrant and possibly pathological splicing patterns.

Moreover, our comprehensive collection of all experimental data available can be helpful also for the design of algorithms for RNA-binding site predictions. Including no-binding sites permits the training of machine learning

algorithms for the prediction of binding sites in a sensitive and specific manner.

Protein interaction data permits to obtain a better understanding of the crosstalk among the splicing machinery and other pathways. These connections, in turn, might highlight the effect of external stimuli or different cellular conditions on splicing regulation.

SpliceAid-F will be continuously updated with quarterly releases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1.

FUNDING

'Ministero dell'Istruzione, Università e Ricerca' (MIUR, Italy): 'PRIN 2009—Large-scale identification and characterization of alternative splicing in the human transcriptome through computational and experimental approaches'; Consiglio Nazionale delle Ricerche: Flagship Project 'Epigen', PNR-CNR Aging Program 2012–2014, Italian Research Foundation for amyotrophic lateral sclerosis (AriSLA) and by Italian Ministry for Foreign Affairs (Italy-Israel actions). Funding for open access charge: Consiglio Nazionale delle Ricerche: Flagship Project ''Epigen''.

Conflict of interest statement. None declared.

REFERENCES

1. Luco,R.F., Allo,M., Schor,I.E., Kornblihtt,A.R. and Misteli,T. (2011) Epigenetics in alternative pre-mRNA splicing. *Cell*, **144**, 16–26.
2. Wahl,M.C., Will,C.L. and Luhrmann,R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
3. Wang,Z. and Burge,C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
4. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
5. Matlin,A.J., Clark,F. and Smith,C.W. (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell. Biol.*, **6**, 386–398.
6. Nilsen,T.W. and Graveley,B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**, 457–463.
7. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
8. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
9. Cartegni,L., Wang,J., Zhu,Z., Zhang,M.Q. and Krainer,A.R. (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.*, **31**, 3568–3571.
10. Huang,H.Y., Chien,C.H., Jen,K.H. and Huang,H.D. (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res.*, **34**, W429–W434.
11. Paz,I., Akerman,M., Dror,I., Kosti,I. and Mandel-Gutfreund,Y. (2010) SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res.*, **38**, W281–W285.
12. Stamm,S., Riethoven,J.J., Le Texier,V., Gopalakrishnan,C., Kumanduri,V., Tang,Y., Barbosa-Morais,N.L. and Thanaraj,T.A. (2006) ASD: a bioinformatics resource on alternative splicing. *Nucleic Acids Res.*, **34**, D46–D55.
13. Raponi,M., Baralle,F.E. and Pagani,F. (2007) Reduced splicing efficiency induced by synonymous substitutions may generate a substrate for natural selection of new splicing isoforms: the case of CFTR exon 12. *Nucleic Acids Res.*, **35**, 606–613.
14. Piva,F., Giulietti,M., Nocchi,L. and Principato,G. (2009) SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics*, **25**, 1211–1213.
15. Piva,F., Giulietti,M., Ballone Burini,A. and Principato,G. (2012) SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum. Mutat.*, **33**, 81–85.
16. Kerrien,S., Aranda,B., Breuza,L., Bridge,A., Broackes-Carter,F., Chen,C., Duesbury,M., Dumousseau,M., Feuermann,M., Hinz,U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, **40**, D841–D846.
17. Kuhn,M., Szklarczyk,D., Franceschini,A., von Mering,C., Jensen,L.J. and Bork,P. (2012) STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Res.*, **40**, D876–D880.
18. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*, **32**, D449–D451.
19. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
20. Stark,C., Breitkreutz,B.J., Chatr-Aryamontri,A., Boucher,L., Oughtred,R., Livstone,M.S., Nixon,J., Van Auken,K., Wang,X., Shi,X. *et al.* (2011) The BioGRID interaction database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.
21. Kalinina,O.V., Wichmann,O., Apic,G. and Russell,R.B. (2012) ProtChemSI: a network of protein-chemical structural interactions. *Nucleic Acids Res.*, **40**, D549–D553.
22. Teplova,M., Song,J., Gaw,H.Y., Teplov,A. and Patel,D.J. (2010) Structural insights into RNA recognition by the alternate-splicing regulator CUG-binding protein 1. *Structure*, **18**, 1364–1377.
23. Valverde,R., Pozdnyakova,I., Kajander,T., Venkatraman,J. and Regan,L. (2007) Fragile X mental retardation syndrome: structure of the KH1-KH2 domains of fragile X mental retardation protein. *Structure*, **15**, 1090–1098.
24. Behzadnia,N., Golas,M.M., Hartmuth,K., Sander,B., Kastner,B., Deckert,J., Dube,P., Will,C.L., Urlaub,H., Stark,H. *et al.* (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.*, **26**, 1737–1748.
25. Du,Z., Lee,J.K., Tjhen,R., Li,S., Pan,H., Stroud,R.M. and James,T.L. (2005) Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 A. *J. Biol. Chem.*, **280**, 38823–38830.
26. Du,Z., Lee,J.K., Fenn,S., Tjhen,R., Stroud,R.M. and James,T.L. (2007) X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *RNA*, **13**, 1043–1051.
27. Schellenberg,M.J., Edwards,R.A., Ritchie,D.B., Kent,O.A., Golas,M.M., Stark,H., Luhrmann,R., Glover,J.N. and MacMillan,A.M. (2006) Crystal structure of a core spliceosomal protein interface. *Proc. Natl Acad. Sci. USA*, **103**, 1266–1271.
28. Sowa,M.E., Bennett,E.J., Gygi,S.P. and Harper,J.W. (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell*, **138**, 389–403.
29. Sanford,J.R., Wang,X., Mort,M., Vanduyn,N., Cooper,D.N., Mooney,S.D., Edenberg,H.J. and Liu,Y. (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.*, **19**, 381–394.
30. Castrignano,T., D'Antonio,M., Anselmo,A., Carrabino,D., D'Onorio De Meo,A., D'Erchia,A.M., Licciulli,F., Mangiulli,M., Mignone,F., Pavesi,G. *et al.* (2008) ASPicDB: a database resource for alternative splicing analysis. *Bioinformatics*, **24**, 1300–1304.

31. Martelli,P.L., D'Antonio,M., Bonizzoni,P., Castrignano,T., D'Erchia,A.M., D'Onorio De Meo,P., Fariselli,P., Finelli,M., Licciulli,F., Mangiulli,M. *et al*. (2011) ASPicDB: a database of annotated transcript and protein variants generated by alternative splicing. *Nucleic Acids Res*., **39**, D80–D85.
32. Gisler,F.M., von Kanel,T., Kraemer,R., Schaller,A. and Gallati,S. (2012) Identification of SNPs in the cystic fibrosis interactome influencing pulmonary progression in cystic fibrosis. *Eur. J. Hum. Genet*, August 15 (doi: 10.1038/ejhg.2012.181; epub ahead of print).
33. Raponi,M., Douglas,A.G., Tammaro,C., Wilson,D.I. and Baralle,D. (2012) Evolutionary constraint helps unmask a splicing regulatory region in BRCA1 exon 11. *PLoS One*, **7**, e37255.
34. Cook,K.B., Kazan,H., Zuberi,K., Morris,Q. and Hughes,T.R. (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res*., **39**, D301–D308.